Proc. Natl. Acad. Sci. USA
Vol. 77, No. 1, pp. 313–317, January 1980
Biochemistry

Organization of the *recA* gene of *Escherichia coli*

(restriction map/nucleotide sequence/amino acid sequence/promoter and terminator of transcription)

TOSHIHIRO HORII, TOMOKO OGAWA, AND HIDEYUKI OGAWA

Department of Biology, Faculty of Science, Osaka University, Toyonaka, Osaka 560, Japan

Communicated by Terrell L. Hill, October 29, 1979

ABSTRACT **The restriction map of a *Bam*HI DNA fragment that contains the *recA* gene of *Escherichia coli* has been established and a large portion of the fragment's nucleotide sequence has been determined. The coding region of the *recA* gene contains 1059 nucleotide residues and encodes a single protein of 353 amino acid residues. The amino acid sequence of the first five residues of the NH_2 terminus of the *recA* protein agrees with the sequence predicted from the DNA sequence except for the absence of formylmethionine in the purified protein. Immediately after the coding sequence, there is a G+C-rich sequence with dyad symmetry followed by an A+T-rich sequence. These could signal termination of transcription. The site of initiation for synthesis *in vitro* of the *recA* messenger RNA has been determined by analysis of the 5′ nucleotide sequence of [γ-^{32}P]ATP-labeled transcripts. The promoter region shows a high degree of symmetry and contains sequences commonly found in recognition and binding sites for RNA polymerase.**

The product of the *recA* gene of *Escherichia coli* plays an essential role in genetic recombination (1), in induction of prophage (2, 3), in ultraviolet light-induced mutagenesis (4), and in repair of various kinds of damage to DNA (1, 5–8). During normal growth this protein is produced at a low level, but when the cell is exposed to any of a number of treatments that damage DNA, an increased rate of production of the protein results (9–13).

At least two gene products seem to regulate expression of the *recA* gene. The *lexA* product is thought to determine the basal rate of expression because mutations in the *lexA* gene result in constitutive expression of the *recA* gene (14). The second regulatory factor is the *recA* protein itself. The *recA* gene of a plasmid that contains only the part of the gene that codes for the NH_2-terminal portion of the protein is expressed at a high rate only in the presence of the active *recA* protein (15). It has been suggested that the *recA* protein inactivates the *lexA* product after DNA damage, thereby causing a higher rate of *recA* expression (11).

Purified *recA* protein can catalyze a number of reactions. These include the hydrolysis of ATP in the presence of single-stranded DNA (15), the ATP-dependent uptake of single-stranded DNA by duplex DNA (16, 17), and the ATP-dependent hybridization of homologous single-stranded DNAs (18). Proteolytic cleavage of the *c*I repressor of phage λ by *recA* protein has also been reported (19).

To facilitate analysis of regulation of expression of the *recA* gene, a set of plasmids that carry the wild-type *recA* gene, a *recA* deletion, or a *recA* mutation has been constructed by *in vitro* gene manipulation techniques (15). One of these plasmids, pTM-2, which contains a wild-type *recA* gene, is cleaved into two fragments by the restriction enzyme *Bam*HI. Analysis of the products of transcription and translation *in vitro* of the smaller fragment has shown that the *recA* gene is present in the fragment (ref. 15; unpublished data). In this paper, we report the restriction enzyme map of the fragment and the nucleotide sequence of the *recA* gene and its neighborhood. The site of initiation of transcription of the *recA* gene *in vitro* has also been determined. The regulation of expression of the *recA* gene is briefly discussed in the light of this structural information.

MATERIALS AND METHODS

Source of the 3-Kilobase (kb) *Bam*HI Fragment. The plasmid pTM-2 was used to prepare a fragment that contains the *recA* gene (15). Cleavage of the plasmid DNA by the restriction enzyme *Bam*HI yielded two fragments, the smaller one 3 kb in size (15).

Digestion with Restriction Endonucleases and Determination of Electrophoretic Mobility of DNA Fragments. The restriction endonucleases *Eco*RI (20), *Bam*HI (21), and *Hga* I (22) were prepared as described. The endonuclease *Ava* II was obtained from Bethesda Research Laboratories, Rockville, MD and *Hin*fI, *Hae* II, *Hae* III, and *Hpa* II were obtained from New England BioLabs. Enzyme digestions were carried out at 37°C in reaction mixtures (50–200 μl) containing 10 mM Tris·HCl (pH 7.6), 8 mM $MgCl_2$, 2 mM 2-mercaptoethanol, and 100 μg of bovine serum albumin per ml, except that 100 mM Tris·HCl (pH 7.6) was used for the digestion with *Eco*RI. Disc gels (0.5 × 12 cm) of 5% or 10% polyacrylamide (acrylamide/bisacrylamide, 19:1) in 36 mM Tris/32 mM KH_2PO_4/1 mM EDTA at pH 7.8 (23) were used for determination of electrophoretic mobility of DNA fragments. The size of a DNA fragment was estimated from its mobility relative to the mobilities of the *Hae* III fragments of ColE1 DNA (J. Tomizawa, personal communication). DNA fragments to be used for further analysis were eluted from crushed gel slices with running buffer and passed through Sephadex G-100.

Transcription of DNA Fragments. The *Hae* III-E fragment from the *Bam*HI fragment was transcribed with *E. coli* RNA polymerase that was prepared as described (24). The 10-μl reaction mixture consisted of 40 mM Tris·HCl at pH 7.9, 100 mM KCl, 10 mM $MgCl_2$, various concentrations of rNTPs including [α-^{32}P]UTP (35 Ci/mmol, Amersham; 1 Ci = 3.7 × 10^{10} becquerels), about 0.03 μg of DNA fragments, and 0.2 unit of RNA polymerase; the mixture was incubated for 30 min at 37°C. For labeling with [γ-^{32}P]ATP (16 Ci/mmol, ICN), 150 μM each of rNTPs, 0.3 μg of DNA, and 2 units of RNA polymerase in 300 μl of reaction mixture were incubated for 30 min at 37°C.

Determination of Nucleotide Sequences. Nucleotide sequences of DNA fragments were determined by the method of Maxam and Gilbert (25). The 5′ ends of DNA fragments were labeled by using polynucleotide kinase. For nucleotide-specific modifications, dimethyl sulfate or hydrazine was used. After cleavage, the products were electrophoresed in a slab gel (0.04–0.1 × 30 × 40 cm) of 15–25% polyacrylamide (acrylamide/bisacrylamide, 19:1) containing 8 M urea, 100 mM Tris borate (pH 8.3), and 2 mM EDTA. The nucleotide sequence

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviation: kb, kilobases.

near the 5′ end of RNA was determined through analysis of the products of partial digestion of [γ-^{32}P]ATP-labeled RNA by RNase A (Worthington), RNase T1, RNase U2 (Sankyo), and RNase Phy I (P-L Biochemicals) essentially as described (26) with some modifications (J. Tomizawa, personal communication).

Sequence Determination of Amino Acids at the NH_2-Terminal Region and Analysis of Amino Acid Composition. The *recA* protein was purified as described (15). The product was greater than 90% pure as determined by sodium dodecyl sulfate/polyacrylamide electrophoresis. The amino acid sequence of the NH_2-terminal region was determined by Edman degradation (27). The amino acid composition of the protein was determined with a Beckman model 120B automated amino acid analyzer as described (28).

RESULTS

Cleavage Map of the 3-kb *Bam*HI Fragments. To construct a cleavage map, the 3-kb *Bam*HI DNA was digested with the restriction endonucleases *Eco*RI, *Hae* II, *Ava* II, *Hae* III, and *Hin*fI. The sizes of the digestion products are presented in Table 1. Two *Eco*RI fragments were further digested with one or more restriction enzymes. The sizes of fragments thus produced are shown in Table 2. The *Ava* II-B, *Hae* II-A, *Hae* III-C, and *Hin*fI-C fragments that were not present in digests of the *Eco*RI-A or -B fragments contain the *Eco*RI site (Fig. 1). Cleavage of the *Eco*RI-A fragment that contained two *Hae* II sites with a combination of the *Hae* II and *Ava* II enzymes did not give the *Ava*II-A fragment, whereas the *Ava* II-D and *Hae*II-B and -C fragments were still generated. Therefore all four *Ava*II DNA fragments can be arranged as shown in Fig. 1. From the results of double digestion of the *Eco*RI-A fragment by the *Hae* II and *Hin*fI enzymes, the *Hin*fI-D and -G fragments, and the *Hae*II-B and -C fragments can be mapped (Table 2 and Fig. 1). By a similar series of double digestion experiments, the locations of the *Hae* III-B and -D and *Hin*fI-F and -H fragments can be assigned by using the *Ava* II and *Hae* II cleavage sites as the key map positions. Also, on the *Eco*RI-B fragment, the *Hae* III and *Hin*fI sites were mapped from the three double digestions (Table 2). Although the *Hae* III-G and -E fragments are located in the *Hin*fI-B fragment, the relative positions of the two cannot be determined from the experiments described above. In addition, small fragments (less than 30 base pairs) could have been missed. Assignment of the precise locations of the cleavage sites is based on the nucleotide sequence, when possible. A map of *Hae* III, *Hin*fI, *Taq* I, and *Hha*I sites in the *Eco*RI-A fragment has recently been reported elsewhere (29).

Table 1. Lengths of fragments produced from the 3-kb *Bam*HI DNA after treatment with restriction endonucleases

Fragment	Fragment length after digestion, base pairs				
	*Eco*RI	*Hae* II	*Ava* II	*Hae* III	*Hin*fI
A	1770*	2600*	1150	890	760
B	1350*	310	1070 (1089)	760	640
C		210	770	570 (599)	640 (652)
D			100 (101)	325 (318)	380
E				300 (302)	275 (280)
F				160 (166)	235 (228)
G				125 (132)	170
H					45 (51)

The 3-kb *Bam*HI DNA was digested with each restriction enzyme. The lengths of the resultant fragments were estimated from their electrophoretic mobilities. The numbers in parentheses are the sizes determined from the nucleotide sequence. For the calculation of the lengths shown here, 6.6 kb is assigned to ColE1 DNA (J. Tomizawa, personal communication). Values indicated by * are calculated as the sum of the lengths of subfragments formed by secondary digestion by other enzymes (see Table 2).

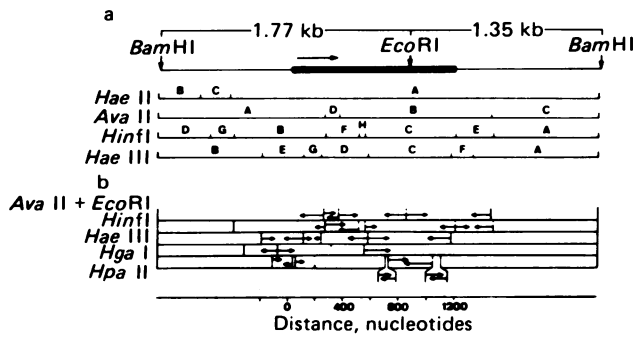

FIG. 1. Cleavage maps and sequencing strategy. (*a*) Cleavage maps of the 3-kb *Bam*HI DNA segment. The map is constructed by using the results presented in Tables 1 and 2 and the information obtained from the nucleotide sequence of Fig. 2. The thickened line indicates the region transcribed *in vitro* by RNA polymerase (15). The horizontal arrow indicates the direction of transcription. (*b*) The fragments used for determination of nucleotide sequence of the region that covers the transcribed region. In addition to the fragments mapped in *a*, some *Hga* I and *Hpa* II fragments were also used. The convenience of the use of these fragments for obtaining information that was difficult to obtain with the fragments mapped in *a* had been known during the course of sequencing the nucleotides. The directions and the extents of the sequence determinations are shown by the arrows. An arrow indicates a 5′-^{32}P-labeled strand aligned in the 5′ to 3′ direction. Numbers on the bottom line indicate the distance from the predominant site for initiation of RNA synthesis *in vitro*.

Nucleotide Sequence of the *recA* Gene. Making use of the restriction enzyme map of the 3-kb *Bam*HI fragment (Fig. 1*a*), we next determined the nucleotide sequence of the middle region of the fragment, which was already thought to be the location of the *recA* gene (15). The DNA fragments used in the sequence determination are indicated in Fig. 1*b*. The sequence is shown in Fig. 2. Inspection of the sequence shows that the region could encode a protein containing 353 amino acids.

Amino Acid Sequence of the NH_2-Terminal Region and the Amino Acid Content of the *recA* Protein. The size of the protein (37,800 daltons) predicted from the nucleotide sequence coincides with that of the *recA* protein (39,000 daltons measured by gel electrophoresis) isolated from bacteria (15). To show that this region in fact encodes the *recA* protein, the amino acid sequence of five residues from the NH_2 terminus of the *recA* protein synthesized *in vivo* (15) was determined (data not shown). The sequence is NH_2-Ala-Ile-Asp-Glu-Asn. This agrees with the sequence predicted from the nucleotide sequence, except for the absence of a formylmethionine residue at the NH_2 terminus. The amino acid composition of the purified *recA* protein also agreed with the composition of the protein predicted from the nucleotide sequence, again with the exception of a formylmethionine residue (data not shown).

Site of Initiation of Transcription. It has been shown that *in vitro* synthesis of the *recA* messenger RNA starts approximately 1 kb to the right of the *Bam*HI site within the *Eco*R1-A segment (15). Inspection of the nucleotide sequence of this region reveals a sequence that has many similarities to sites known to act as promoters for *E. coli* RNA polymerase (discussion below). Considering this information, we determined a site of

Table 2. Lengths of the fragments produced from the *Eco*RI-A and *Eco*RI-B fragments after treatment with one or more restriction enzymes

	*Eco*RI-A fragment									*Eco*RI-B fragment					
	Ava II	*Hae* II	*Hae* III	*Hin*fI	*Ava* II + *Hae* II	*Hin*fI + *Hae* II	*Hae* III + *Ava* II	*Hin*fI + *Ava* II	*Hae* III + *Hin*fI	*Ava* II	*Hae* III	*Hin*fI	*Hae* III + *Ava* II	*Hin*fI + *Ava* II	*Hae* III + *Hin*fI
Lengths of fragments with heterologous ends,* base pairs	470†	1180†	280†	290†	600 470†	290† 140 70	280† 180	290† 130 90	280† 240	600†	330†	360†	330† 125	360† 250	330† 135 115 35
Fragments with homologous ends	AII-A AII-D	HII-B HII-C	HIII-B HIII-D HIII-E HIII-G	FI-B FI-D FI-F FI-G FI-H	HII-B HII-C AII-D	FI-B FI-F FI-H HII-B	HIII-B HIII-E HIII-G AII-D	FI-B FI-D FI-G FI-H	HIII-E HIII-G FI-D FI-F FI-G FI-H	AII-C	HIII-A HIII-F	FI-A FI-E	HIII-F AII-C	FI-A	FI-A

The *Eco*RI-A and *Eco*RI-B fragments were digested with one restriction enzyme or various pairs of them and the lengths of the fragments formed were determined. Each fragment is identified by a combination of the symbols under the number of base pairs. Symbols RI, AII, HII, HIII, and FI represent the restriction enzymes *Eco*RI, *Ava* II, *Hae* II, *Hae* III, and *Hin*fI, respectively, and, for example, AII-A means *Ava* II-A fragment.

* Fragments with a *Bam*HI end are included with fragments with homologous ends.

† Fragments with an *Eco*RI end.

initiation of transcription in the *Hae*III-E fragment that contains the suspected region.

RNA was synthesized on the fragment with RNA polymerase. In a reaction mixture the concentration of each rNTP was 10 μM including [α-^{32}P]UTP. Three other reaction mixtures contained 200 μM each of ATP, GTP, or CTP and other rNTPs at 10 μM each. Electrophoretic analysis of the products from these reactions showed that a single species of RNA of approximately 95 nucleotides was always the predominant product (Fig. 3). An increase in the ATP concentration stimulated RNA synthesis about 5-fold, whereas increases in the GTP or CTP concentrations had small stimulatory or inhibitory effects, respectively. These results suggest that the transcription started mostly with ATP, and set the site of transcription initiation at around the position shown as no. 1 in Fig. 2.

The exact position of initiation was determined by the analysis of the nucleotide sequence at the 5′ end of the RNA synthesized *in vitro*. The sequence was deduced from the products of partial digestion (26) of the [γ-^{32}P]ATP-labeled RNA by various RNases. The result in Fig. 4 shows that the 5′ sequence of the major labeled RNA is pppA-A-C-A-G-A-A-C-A-U-A-. The presence of faint bands such as di-, tri-, and

```
                                -100                                                -50                                                +1
TGCTTCAGCGGCGACCGTGATGCGGTGCGTCGTCAGGCTACTGCGTATGCTTGCAGACCTTGTGGCAACAATTTCTACAAAACACTTGATACTGTATGAGCATACAGTATAATTGCTTCAACAGAACATATT

                                      +50                                                +100
GACTATCCGGTATTACCCGGCATGACAGGAGTAAAAATGGCTATCGACGAAAACAAACAGAAAGCGTTGGCGGCAGCACTGGGCCAGATTGAGAAACAATTTGGTAAAGGCTCCATCATGCGCCTGGGTGAA
                                      fMetAlaIleAspGluAsnLysGlnLysAlaLeuAlaAlaAlaLeuGlyGlnIleGluLysGlnPheGlyLysGlySerIleMetArgLeuGlyGlu
   +150                                                +200                                                +250
GACCGTTCCATGGATGTGGAAACCATCTCTACCGGTTCGCTTTCACTGGATATCGCGCTTGGGGCAGGTGGTCTGCCGATGGGCCGTATCGTCGAAATCTACGGACCGGAATCTTCCGGTAAAACCACGCTG
AspArgSerMetAspValGluThrIleSerThrGlySerLeuSerLeuAspIleAlaLeuGlyAlaGlyGlyLeuProMetGlyArgIleValGluIleTyrGlyProGluSerSerGlyLysThrThrLeu
                         +300                                                +350                                                +400
ACGCTGCAGGTGATCGCCGCAGCGCAGCGTGAAGGTAAAACCTGTGCGTTTATCGATGCTGAACACGCGCTGGACCCAATCTACGCACGTAAACTGGGCGTCGATATCGACAACCTGCTGTGCTCCCAGCCG
ThrLeuGlnValIleAlaAlaAlaGlnArgGluGlyLysThrCysAlaPheIleAspAlaGluHisAlaLeuAspProIleTyrAlaArgLysLeuGlyValAspIleAspAsnLeuLeuCysSerGlnPro
                                      +450                                                +500
GACACCGGCGAGCAGGCACTGGAAATCTGTGACGCCCTGGCGCGTTCTGGCGCAGTAGACGTTATCGTCGTTGACTCCGTGGCGGCACTGACGCCGAAAGCGGAAATCGAAGGCGAAATCGGCGACTCTCAC
AspThrGlyGluGlnAlaLeuGluIleCysAspAlaLeuAlaArgSerGlyAlaValAspValIleValValAspSerValAlaAlaLeuThrProLysAlaGluIleGluGlyGluIleGlyAspSerHis
         +550                                                +600                                                +650
ATGGGCCTTGCGGCACGTATGATGAGCCAGGCGATGCGTAAGCTGGCGGGTAACCTGAAGCAGTCCAACACGCTGCTGATCTTCATCAACCAGATCCGTATGAAAATTGGTGTGATGTTCGGTAACCCGGAA
MetGlyLeuAlaAlaArgMetMetSerGlnAlaMetArgLysLeuAlaGlyAsnLeuLysGlnSerAsnThrLeuLeuIlePheIleAsnGlnIleArgMetLysIleGlyValMetPheGlyAsnProGlu
                         +700                                                +750                                                +800
ACCACTACCGGTGGTAACGCGCTGAAATTCTACGCCTCTGTTCGTCTCGACATCCGTCGTATCGGCGCGGTGAAAGAGGGCGAAAACGTGGTGGGTAGCGAAACCCGCGTGAAAGTGGTGAAGAACAAAATC
ThrThrThrGlyGlyAsnAlaLeuLysPheTyrAlaSerValArgLeuAspIleArgArgIleGlyAlaValLysGluGlyGluAsnValValGlySerGluThrArgValLysValValLysAsnLysIle
                                      +850                                                +900
GCTGCGCCGTTTAAACAGGCTGAATTCCAGATCCTCTACGGCGAAGGTATCAACTTCTACGGCGAACTGGTTGACCTGGGCGTAAAAGAGAAGCTGATCGAGAAAGCAGGCGCGTGGTACAGCTACAAAGGT
AlaAlaProPheLysGlnAlaGluPheGlnIleLeuTyrGlyGluGlyIleAsnPheTyrGlyGluLeuValAspLeuGlyValLysGluLysLeuIleGluLysAlaGlyAlaTrpTyrSerTyrLysGly
         +950                                                +1000                                                +1050
GAGAAGATCGGTCAGGGTAAAGCGAATGCGACTGCCTGGCTGAAAGATAACCCGGAAACCGCGAAAGAGATCGAGAAGAAAGTACGTGAGTTGCTGCTGAGCAACCCGAACTCAACGCCGGATTTCTCTGTA
GluLysIleGlyGlnGlyLysAlaAsnAlaThrAlaTrpLeuLysAspAsnProGluThrAlaLysGluIleGluLysLysValArgGluLeuLeuLeuSerAsnProAsnSerThrProAspPheSerVal
                                      +1100                                                +1150                                                +1200
GATGATAGCGAAGGCGTAGCAGAAACTAACGAAGATTTTTAATCGTCTTGTTTGATACACAAGGGCGCATCTGCGGCCCTTTTGCTTTTTTAAGTTGTAAGGATATGCCATGACAGAATCAACATCCCGTCG
AspAspSerGluGlyValAlaGluThrAsnGluAspPhe
```

FIG. 2. Nucleotide sequence of the *recA* gene and its neighborhood, and the amino acid sequence of the encoded protein. The DNA sequence is numbered beginning at the predominant site of initiation of transcription *in vitro*.

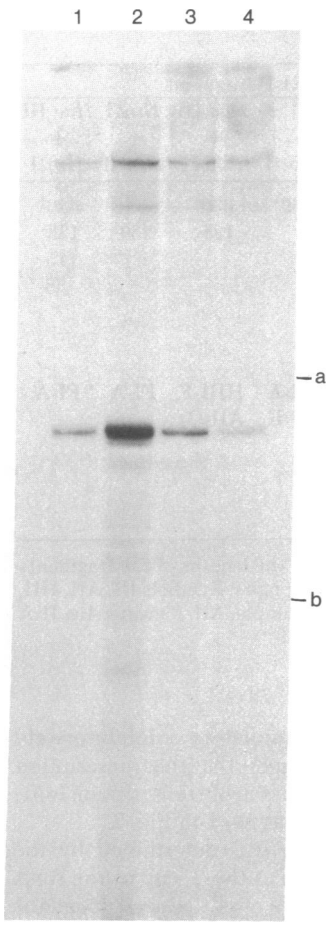

FIG. 3. Autoradiograph of transcripts synthesized from the *Hae*III-E fragments by RNA polymerase. Transcripts were labeled with [α-^{32}P]UTP in the presence of 10 μM of each rNTP, except that the concentrations of ATP, GTP, and CTP were 200 μM for transcripts shown in lane 2, 3, and 4, respectively. Electrophoresis in an 8% acrylamide/urea gel was carried out at 1000 V for 3 hr. The positions indicated by a and b are location of bands formed by RNA-1 of ColE1 [108 nucleotides, (30)] and 4S RNA of phage λ [77 nucleotides (31)].

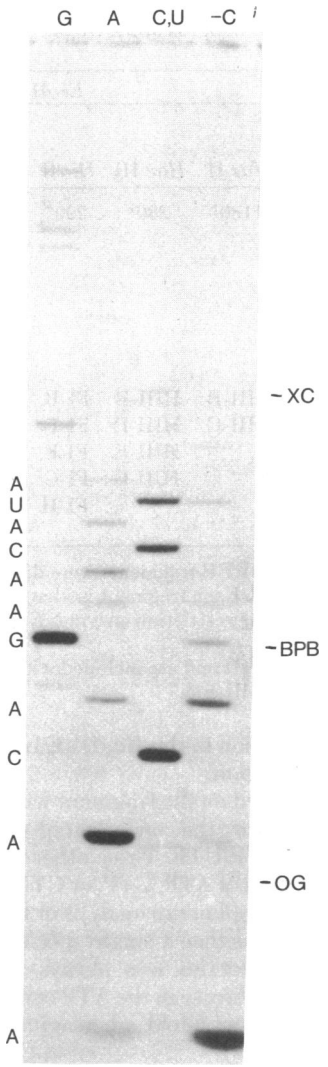

FIG. 4. Autoradiograph of partial digests by various RNases of [γ-^{32}P]ATP-labeled nucleotides synthesized from the *Hae*III-E fragments. [γ-^{32}P]ATP-labeled transcripts from the *Hae*III-E fragments were fractionated in an 8% gel as described in Fig. 3 and RNA about 95 nucleotides long was extracted electrophoretically and concentrated by ethanol precipitation in the presence of 20 μg of *E. coli* tRNA. The precipitate was dissolved in water, divided into four equal portions, and dried. The sample was dissolved in 4 μl of the buffer to which 1 μl of an RNase solution was added. The conditions used (T. Itoh and J. Tomizawa, personal communication) were as follows: lane G, 0.03 μg of RNase T1 in 0.1 M Tris·HCl/10 mM EDTA, pH 7.5; lane A, 0.01 unit of RNase U2 in 20 μM sodium citrate, pH 3.5; lane C,U, 3 ng of RNase A in the same buffer as for RNase T1 digestion; and lane −C, 0.5 unit of RNase Phy I in 10 mM sodium acetate/1 mM EDTA, pH 5.0. Samples were incubated for 30 min at 0°C except that the reaction with RNase Phy I was carried out at 25°C. Gel electrophoresis was carried out in 25% acrylamide/urea gel at 2000 V for 5 hr. XC, BPB, and OG indicate the positions of xylene cyanol FF, bromphenol blue, and orange G, respectively. It has been shown that orange G moves slightly faster than dinucleoside pentaphosphate and the mobility of bromphenol blue is similar to that of hexanucleoside nonaphosphate (J. Tomizawa, personal communication).

tetranucleotides formed by digestion with RNase A, RNase U2, and RNase T1, respectively, suggests the existence of a minor transcript that begins at a position corresponding to the A residue next to the major initiation site. Therefore, transcription initiates most frequently at position 1 and much less frequently at position 2 in the nucleotide sequence (Fig. 2).

DISCUSSION

Initiation of Transcription. The known promoter regions for *E. coli* RNA polymerase have certain structural homologies, particularly in two regions about 10 and 35 nucleotides upstream of the site where transcription begins. The latter probably provides a site recognized by the RNA polymerase, whereas the former is where the polymerase binds (32). Approximately 10 nucleotides upstream of the position where transcription of the *recA* gene begins, there is the sequence T-A-T-A-A-T-T, which matches the general structure of the RNA polymerase binding site (33, 34). The same sequence is present in the C_{17} promoter of the phage λ (35). About 20 nucleotides further upstream, a sequence that includes the highly common T-T-G-A and surrounding sequence exists in the recognition region. These arrangements of nucleotides frequently found in promoter regions are indicated in Fig. 5.

Termination of Transcription. It has been shown that transcription *in vitro* of the 3-kb *Bam*HI fragment frequently terminates after synthesis of RNA of about 1 kb (ref. 15; unpublished data). In the nucleotide sequence of the suspected region of transcription termination, there is a T-rich stretch preceded by two sequences with dyad symmetry. This region probably signals termination of transcription (32, 33, 36). Of the two symmetrical sequences, the one closer to the T-rich stretch is larger and is richer in G·C pairs. The transcription probably terminates in the T-rich stretch (36).

Translation. About 50 nucleotides downstream from the initiation site of transcription, there is an AUG codon that begins the structural portion of the *recA* gene. About 10 nucleotides upstream from the initiation codon, there is the sequence A-G-G-A-G, which could serve as a ribosome binding site (37). Starting at this AUG, a protein of 353 amino acids would be

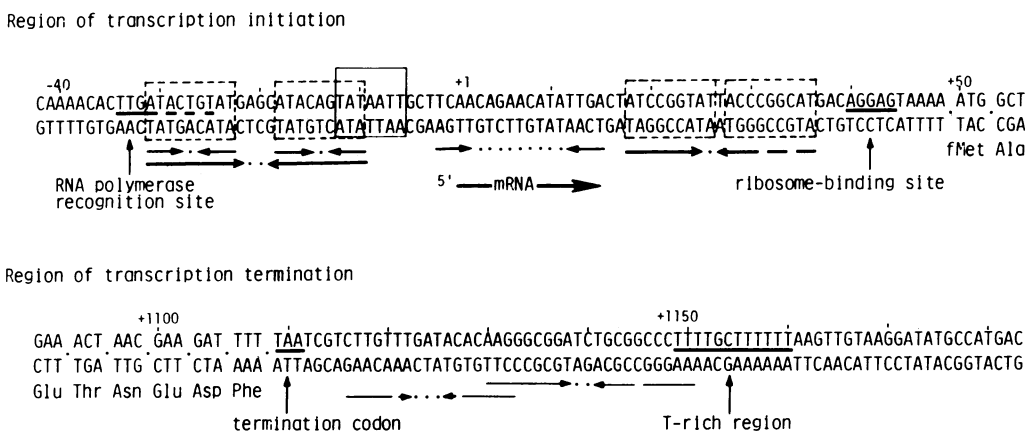

FIG. 5. Nucleotide sequence of the regions of initiation and termination of transcription. The heptanucleotide sequence that probably serves as the binding site for the RNA polymerase is boxed with solid lines. The presumed RNA polymerase recognition site, the ribosome-binding site, the termination codon, and the T-rich sequence in the region of transcription termination are underlined. The regions of dyad symmetries are indicated by arrows, and homologous dyad symmetries are boxed with broken lines. Some amino acid residues encoded in the regions are also presented.

made. No protein containing more than 53 amino acid residues could be made by reading the transcript in a different frame.

Regulation of Expression of the *recA* Gene. The coding sequence of the *recA* protein is surrounded by the promoter and terminator of transcription. Therefore, the *recA* gene is probably monocistronic. While the *recA* protein is produced at a low basal level during normal cell growth, it is formed more efficiently after various treatments that cause DNA damage. These observations indicate the presence of a mechanism that regulates expression of the *recA* gene. It has been suggested that the *lexA* protein inhibits the expression of the *recA* gene (38). The complex structure with dyad symmetries in the promoter region (Fig. 5) could contain the site where the *lexA* protein interacts with the DNA.

We thank Drs. T. Itoh and J. Tomizawa for collaboration in determination of the RNA sequence, Drs. M. Takanami and K. Sugimoto for introducing us to the DNA nucleotide sequencing technique, Mr. T. Hase and Dr. H. Matsubara for assistance in determination of the amino acid sequence, and Dr. G. Selzer for critical reading of the manuscript. This work was supported by a grant-in-aid for Scientific Research and for Cancer Research from the Ministry of Education, Science, and Culture of Japan.

1. Clark, A. J. & Margulies, A. D. (1965) *Proc. Natl. Acad. Sci. USA* **53,** 451–459.
2. Brooks, K. & Clark, A. J. (1967) *J. Virol.* **1,** 283–293.
3. Hertman, I. & Luria, S. E. (1967) *J. Mol. Biol.* **23,** 117–133.
4. Miura, A. & Tomizawa, J. (1968) *Mol. Gen. Genet.* **103,** 1–10.
5. Howard-Flanders, P. & Boyce, R. P. (1966) *Radiat. Res. Suppl.* **6,** 156–184.
6. Tomizawa, J. & Ogawa, H. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **33,** 243–251.
7. Jenkins, S. T. & Bennett, P. M. (1976) *J. Bacteriol.* **125,** 1214–1216.
8. Witkin, E. M. (1976) *Bacteriol. Rev.* **40,** 869–907.
9. Gudas, L. J. & Pardee, A. B. (1976) *J. Mol. Biol.* **101,** 459–477.
10. McEntee, K. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5275–5279.
11. Gudas, L. J. & Mount, D. W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5280–5284.
12. Emmerson, P. T. & West, S. C. (1977) *Mol. Gen. Genet.* **155,** 77–85.
13. Little, J. W. & Kleid, D. G. (1977) *J. Biol. Chem.* **252,** 6251–6252.
14. Mount, D. W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 300–304.
15. Ogawa, T., Wabiko, H., Tsujimoto, T., Horii, T., Masukata, H. & Ogawa, H. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43,** 909–915.
16. Shibata, T., DasGupta, C., Cunningham, R. P. & Radding, C. M. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 1638–1642.
17. McEntee, K., Weinstock, G. M. & Lehman, I. R. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 2615–2619.
18. Weinstock, G. M., McEntee, K. & Lehman, I. R. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 126–130.
19. Roberts, J. W., Roberts, C. W. & Craig, N. L. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 4714–4718.
20. Yoshimori, R. N. (1971) Dissertation (Univ. of California, San Francisco, CA).
21. Wilson, G. A. & Young, F. E. (1975) *J. Mol. Biol.* **97,** 123–125.
22. Roberts, R. J., Breitmeyer, J. B., Takachinik, N. F. & Meyers, P. A. (1975) *J. Mol. Biol.* **91,** 121–123.
23. Oka, A. (1978) *J. Bacteriol.* **133,** 916–924.
24. Burgess, R. R. & Jendrisak, J. J. (1975) *Biochemistry* **14,** 4634–4642.
25. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
26. Simoncsitz, A., Brownlee, G. G., Brown, R. S., Rubin, J. R. Guilley, H. (1977) *Nature (London)* **269,** 833–836.
27. Edman, P. (1970) in *Molecular Biology, Biochemistry and Biophysics,* eds. Kleinzeller, A., Springer, G. F. & Wittman, H. G. (Springer-Verlag, Berlin), Vol. 8, pp. 211–266.
28. Spackman, D. H., Stein, W. H. & Moore, S. (1958) *Anal. Chem.* **30,** 1190–1206.
29. Sancar, A. & Rupp, W. D. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 3144–3148.
30. Morita, M. & Oka, A. (1979) *Eur. J. Biochem.* **97,** 435–443.
31. Rosenberg, M., de Crombrugghe, B. & Musso, R. (1976) *Proc. Natl. Acad. Sci. USA* **76,** 717–721.
32. Gilbert, W. (1976) in *RNA Polymerase,* eds. Losick, R. & Chamberlin, M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 193–205.
33. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 784–788.
34. Schaller, H., Gray, C. & Herrmann K. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 737–741.
35. Rosenberg, M., Court, D., Shimatake, H., Brady, C. & Wulff, D. L. (1978) *Nature (London)* **272,** 414–422.
36. Adhya, S. & Gottesman, M. (1978) *Annu. Rev. Biochem.* **47,** 967–996.
37. Shine, J. & Dalgarno, L. (1975) *Nature (London)* **254,** 34–38.
38. Gudas, L. J. (1976) *J. Mol. Biol.* **104,** 567–587.